Sur l'approximation des lois de lissage marginales par les algorithmes two-filter

Sylvain Le Corff [1] & Thi Ngoc Minh Nguyen[2] & Éric Moulines [3]

[1] *Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay. sylvain.lecorff@math.u-psud.fr*
[2] *LTCI, CNRS and Télécom ParisTech. thi-ngoc-minh.nguyen@telecom-paristech.fr*
[3] *Centre de Mathématiques Appliquées, Ecole Polytechnique. eric.moulines@polytechnique.edu*

Résumé. Cette présentation s'intéresse à l'estimation de lois a posteriori dans les modèles de Markov cachés. L'approximation des lois (dites de lissage) de certains états cachés conditionnellement à toutes les observations (passées, présentes, futures) est un problème délicat bien que crucial pour l'estimation d'états cachés ou de paramètres au sens du maximum de vraisemblance. Nous proposons une analyse des approximations de ces lois fournies par les méthodes two-filter. Ces méthodes combinent l'utilisation d'un filtre particulaire approchant les lois de filtrage à celle d'un filtre backward approchant une quantité proportionnelle à la loi d'un état conditionnellement aux observations futures. Nous établissons des inégalités de déviation exponentielles ainsi qu'un théorème central limite, ce qui permet d'avoir pour ces algorithmes des résultats du même ordre que ceux existants pour d'autres algorithmes particulaires. L'un des intérêts majeurs de ces résultats est la forme particulière de la variance asymptotique qui peut-être estimée en ligne pour les algorithmes two-filter, seuls algorithmes pour lesquels cela est possible à ce jour.

Mots-clés. Modèles de Markov cachés; Méthodes de Monte Carlo Séquentielles; Lissage; Two-filter.

Abstract. A prevalent problem in general state space models is the approximation of the smoothing distribution of a state conditional on the observations from the past, the present, and the future. The aim of this talk is to provide a rigorous analysis of such approximations of smoothing distributions provided by the two-filter algorithms. These methods combine a forward filter approximating the filtering distributions with a backward information filter approximating a quantity proportional to the posterior distribution of the state given future observations. Nonasymptotic deviation inequalities and central limit theorems are established for the estimation of the marginal smoothing distributions using the several two-filter algorithms proposed in the literature. One of the main consequences of the central limit theorem is that the asymptotic variance is the sum of the forward filter asymptotic variance and the backward information filter asymptotic variance. These two quantities may be consistently estimated on-the-fly to quantify Monte Carlo errors.

Keywords. Hidden Markov models; Smoothing; Sequential Monte Carlo methods; Two-filter.

Algorithmes two-filter

Dans cette présentation, nous considérons un modèle de Markov caché, c'est-à-dire un processus $\{(X_t, Y_t)\}_{t\geq 0}$ tel que la chaîne de Markov $\{X_t\}_{t\geq 0}$ n'est observée qu'au travers des observations $\{Y_t\}_{t\geq 0}$. Ces modèles de Markov à données cachées jouent un rôle fondamental dans de nombreuses disciplines : ingénierie, traitement du signal, poursuite de cible, identification de systèmes complexes, biologie, voir par exemple [5, 4].

L'estimation de tels modèles requiert la connaissance des lois (dites de lissage) de suites d'états cachés conditionnellement à toutes les observations $(Y_0, \dots, Y_T)$ (dans le but d'utiliser par exemple des algorithmes du type Expectation Maximization ou des méthodes de gradient stochastiques). Cependant, ces lois ne peuvent être calculées explicitement dans les modèles généraux et nous nous intéressons à une classe particulière de méthodes de Monte Carlo séquentielles, appelées algorithmes two-filter, pour les approcher. Ces approximations sont calculées à l'aide d'échantillons aléatoires, les particules, associées à des poids d'importance. Nous proposons d'étendre les résultats de convergence connus aux algorithmes two-filter. Ces méthodes introduites par [9] puis développées dans [1] et [7] combinent deux filtres particulaires indépendants, un premier évoluant de $t = 0$ à $t = T$ approchant les lois de filtrage (loi de X_t conditionnellement à $Y_{0:t}$) et un second évoluant de $t = T$ à $t = 0$ approchant une quantité proportionnelle à la loi de X_t conditionnellement aux observations futures $Y_{t:T}$.

Considérons $(\mathsf{X}, \mathcal{X})$ et $(\mathsf{Y}, \mathcal{Y})$ deux espaces mesurables, Q un noyau de Markov défini sur $\mathsf{X} \times \mathcal{X}$ et $\{g_t\}_{t\geq 0}$ une famille de fonctions positives sur X. Pour tout $x \in \mathsf{X}$, $Q(x, \cdot)$ a pour densité $q(x, \cdot)$ par rapport à une mesure λ sur $(\mathsf{X}, \mathcal{X})$. Pour toute fonction h sur X, toute distribution de probabilité χ sur $(\mathsf{X}, \mathcal{X})$ et tout $0 \leq s \leq t \leq T$, considérons la distribution de lissage marginale $\phi_{\chi,s|T}$ définie par :

$$\phi_{\chi,s|T}[h] := \frac{\int \chi(\mathrm{d}x_0) g_0(x_0) \prod_{u=1}^{T} Q(x_{u-1}, \mathrm{d}x_u) g_u(x_u) h(x_s)}{\int \chi(\mathrm{d}x_0) g_0(x_0) \prod_{u=1}^{T} Q(x_{u-1}, \mathrm{d}x_u) g_u(x_u)} \; . \tag{1}$$

Dans le cas où $\{X_t\}_{t\geq 0}$ est une chaîne de Markov de noyau de transition Q et de loi initiale χ et où les observations $\{Y_t\}_{t\geq 0}$ sont indépendantes conditionnellement à $\{X_t\}_{t\geq 0}$ et telles que pour tout $k \geq 0$ la loi conditionnelle de Y_k a pour densité $g_k(X_k, \cdot)$ (avec par convention $g_k(X_k, Y_k) = g_k(X_k)$), (1) peut s'interpréter de la façon suivante :

$$\phi_{\chi,s|T}[h] = \mathbb{E}\left[h(X_s) | Y_{0:T}\right] \; .$$

Les distributions de filtrage sont données par $\phi_{\chi,s} = \phi_{\chi,s|s}$, ce qui s'écrit dans ce cas :

$$\phi_{\chi,s}[h] = \mathbb{E}\left[h(X_s)|Y_{0:s}\right] \ .$$

Filtre forward

Les filtres particulaires introduits par [8, 9] permettent dans un premier temps d'approcher les lois $\phi_{\chi,s}$ pour $0 \leq s \leq T$. Ces méhodes propagent une suite de particules $\{\xi_s^\ell\}_{\ell=1}^N$ associées à des poids d'importance $\{\omega_s^\ell\}_{\ell=1}^N$ utilisant des étapes successives d'échantillonnage d'importance et de réechantillonnage permettant de dupliquer les particules ayant des poids importants et de supprimer celles ayant un poids négligeable. Nous détaillerons le mécanisme permettant de produire récursivement (au fur et à mesure que les observations sont reçues) ces particules et leurs poids. Ainsi, à chaque instant, nous obtenons l'approximation suivante de (1) :

$$\phi_{\chi,s}^N[h] := \Omega_s^{-1} \sum_{\ell=1}^N \omega_s^\ell h(\xi_s^\ell) \ , \quad \text{où} \quad \Omega_s := \sum_{\ell=1}^N \omega_s^\ell \ . \tag{2}$$

Filtre backward

Considérons ensuite une famille $\{\gamma_t\}_{t\geq 0}$ de fonctions mesurables positives sur X. En s'inspirant de [1], pour tout $0 \leq t \leq T$ on peut introduire le filtre d'information backward $\psi_{\gamma,t|T}$ défini, pour toute function mesurable h, par :

$$\psi_{\gamma,t|T}[h] := \frac{\int \gamma_t(x_t)\,\mathrm{d}x_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1})\,Q(x_{u-1},\mathrm{d}x_u)\right] g_T(x_T)h(x_t)}{\int \gamma_t(x_t)\,\mathrm{d}x_t \left[\prod_{u=t+1}^T g_{u-1}(x_{u-1})\,Q(x_{u-1},\mathrm{d}x_u)\right] g_T(x_T)} \ .$$

Si X_t a pour densité de probabilité γ_t, alors $\psi_{\gamma,t|T}$ est la loi conditionnelle de X_t sachant $Y_{t:T}$. Le filtre $\psi_{\gamma,t|T}$ pouvant se calculer de façon récursive de $t = T$ à $t = 0$, il est également possible d'approcher à chaque instant $\psi_{\gamma,t|T}[h]$ par une suite de particules associées à des poids $\{(\breve{\xi}_{t|T}^i, \breve{\omega}_{t|T}^i)\}_{i=1}^N$ en suivant la même démarche que pour le filtre forward :

$$\psi_{\gamma,t|T}^N[h] := \breve{\Omega}_{t|T}^{-1} \sum_{i=1}^N \breve{\omega}_{t|T}^i h(\breve{\xi}_{t|T}^i) \ , \quad \text{où} \quad \breve{\Omega}_{t|T} := \sum_{i=1}^N \breve{\omega}_{t|T}^i \ . \tag{3}$$

Approximation des lois de lissage marginales

En utilisant les définitions des filtres forward et backward, la loi de lissage marginale peut se décomposer de la façon suivante :

$$\phi_{\chi,s|T}[h] \propto \int \phi_{\chi,s-1}(\mathrm{d}x_{s-1})\psi_{\gamma,s+1|T}(\mathrm{d}x_{s+1})q(x_{s-1},x_s)g_s(x_s)\frac{q(x_s,x_{s+1})}{\gamma_{s+1}(x_{s+1})}h(x_s)\mathrm{d}x_s \ . \tag{4}$$

En remplaçant les filtres dans cette expression par leurs approximations (2) et (3), [7] et [1] ont proposé différentes approches qui seront détaillées lors de la présentation, pour estimer $\phi_{\chi,s|T}[h]$ à l'aide de particules et de poids $\{(\tilde{\xi}^i_{s|T}, \tilde{\omega}^i_{s|T})\}_{i=1}^N$:

$$\phi^N_{\chi,s|T}[h] = \sum_{i=1}^N \frac{\tilde{\omega}^i_{s|T}}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}^i_{s|T}) \ .$$

Résultats de convergence

Nous montrerons que l'échantillon $\{(\tilde{\omega}^i_{s|T}, \tilde{\xi}^\ell_{s|T})\}_{\ell=1}^N$ fournit une approximation convergente de $\phi_{\chi,s|T}$. Sous des hypothèses faibles sur le modèle et sur les mécanismes utilisés nous prouvons que pour tout $s < T$, il existe $0 < B_{s|T}, C_{s|T} < \infty$ tels que pour tout $N \geq 1$, $\varepsilon > 0$ et toute fonction h sur X,

$$\mathbb{P}\left(\left|\sum_{i=1}^N \frac{\tilde{\omega}^i_{s|T}}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}^i_{s|T}) - \phi_{\chi,s|T}[h]\right| > \epsilon\right) \leq B_{s|T} \mathrm{e}^{-C_{s|T} N \epsilon^2} \ . \tag{5}$$

Nous montrerons également que sous des hypothèses plus fortes, en supposant par exemple que le modèle considéré est fortement mélangeant, les constantes dans la borne précédente sont uniformes en temps. L'inégalité (5) permet d'obtenir un résultat de convergence du même ordre que celui des approximations obtenues par d'autres méthodes particulaires telles que celles données dans [2, 3, 6]. Cependant, l'un des intérêts de la méthode présentée ici est qu'elle peut être implémentée avec une complexité linéaire en fonction de N sans hypothèse sur le modèle.

L'autre résultat plus important que nous présenterons est un théorème central limite. Pour tout $1 \leq s \leq T-1$ il existe une fonction $\Upsilon_{\chi,s|T}$ telle que pour toute fonction h,

$$\sqrt{N}\left(\sum_{i=1}^N \frac{\tilde{\omega}^i_{s|T}}{\tilde{\Omega}_{s|T}} h(\tilde{\xi}^i_{s|T}) - \phi_{\chi,s|T}[h]\right) \xrightarrow{\mathcal{D}}_{N\to\infty} \mathcal{N}\left(0, \Upsilon_{\chi,s|T}\left[h\right]\right) \ .$$

Une fois encore, sous des hypothèses de mélange fort, la variance asymptotique est uniformément bornée en temps ce qui est conforme avec les résultats connus pour les autres méthodes de Monte Carlo séquentielles utilisées pour l'approximation des lois de lissage marginales. Cependant, les méthodes de type two-filter reposant sur deux filtres particulaires indépendants, la fonction $\Upsilon_{\chi,s|T}$ a une structure très particulière. Cette fonction est la somme de deux contributions qui sont des variances asymptotiques de filtres particulaires. Grâce aux travaux récents de [10] cela nous permet de proposer un estimateur de la variance asymptotique $\Upsilon_{\chi,s|T}\left[h\right]$ en ligne utilisant uniquement les variables simulées par les deux filtres. Nous pouvons ainsi quantifier les erreurs Monte Carlo de notre méthode d'estimation ce qui n'est à ce jour possible pour aucune autre méthode de lissage.

Tous ces résultats seront illustrés à l'aide de simulations numériques utilisant des données issues de différents modèles.

References

[1] M. Briers, A. Doucet, and S. Maskell. Smoothing algorithms for state-space models. *Annals Institute Statistical Mathematics*, 62(1):61–89, 2010.

[2] P. Del Moral, A. Doucet, and S. Singh. A Backward Particle Interpretation of Feynman-Kac Formulae. *ESAIM M2AN*, 44(5):947–975, 2010.

[3] R. Douc, A. Garivier, E. Moulines, and J. Olsson. Sequential Monte Carlo smoothing for general state space hidden Markov models. *Ann. Appl. Probab.*, 21(6):2109–2145, 2011.

[4] R. Douc, E. Moulines, and D.S. Stoffer. *Nonlinear time series: theory, methods and applications with R examples.* CRC Press, 2014.

[5] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Springer, 2001.

[6] C. Dubarry and S. Le Corff. Non-asymptotic deviation inequalities for smoothed additive functionals in nonlinear state-space models. *Bernoulli*, 19(5B):2222–2249, 2013.

[7] P. Fearnhead, D. Wyncoll, and J. Tawn. A sequential smoothing algorithm with linear computational cost. *Biometrika*, 97(2):447–464, 2010.

[8] N. Gordon, D. Salmond, and A.F. Smith. Novel approach to nonlinear/non-Gaussian bayesian state estimation. *IEE Proc. F, Radar Signal Process*, 140:107–113, 1993.

[9] G. Kitagawa. Monte-Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, 1:1–25, 1996.

[10] A. Lee and N. Whiteley. Variance estimation and allocation in the particle filter. *arXiv:1509.00394*, 2015.